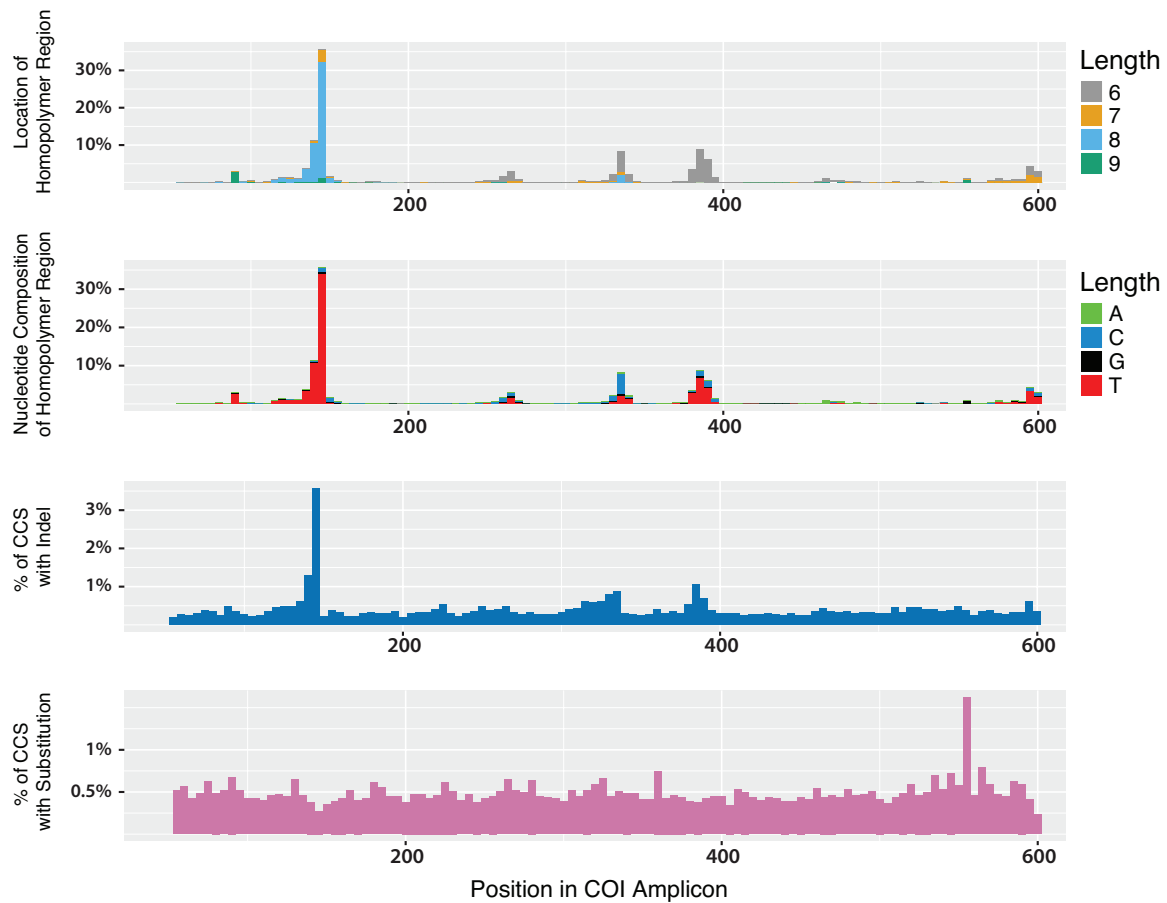
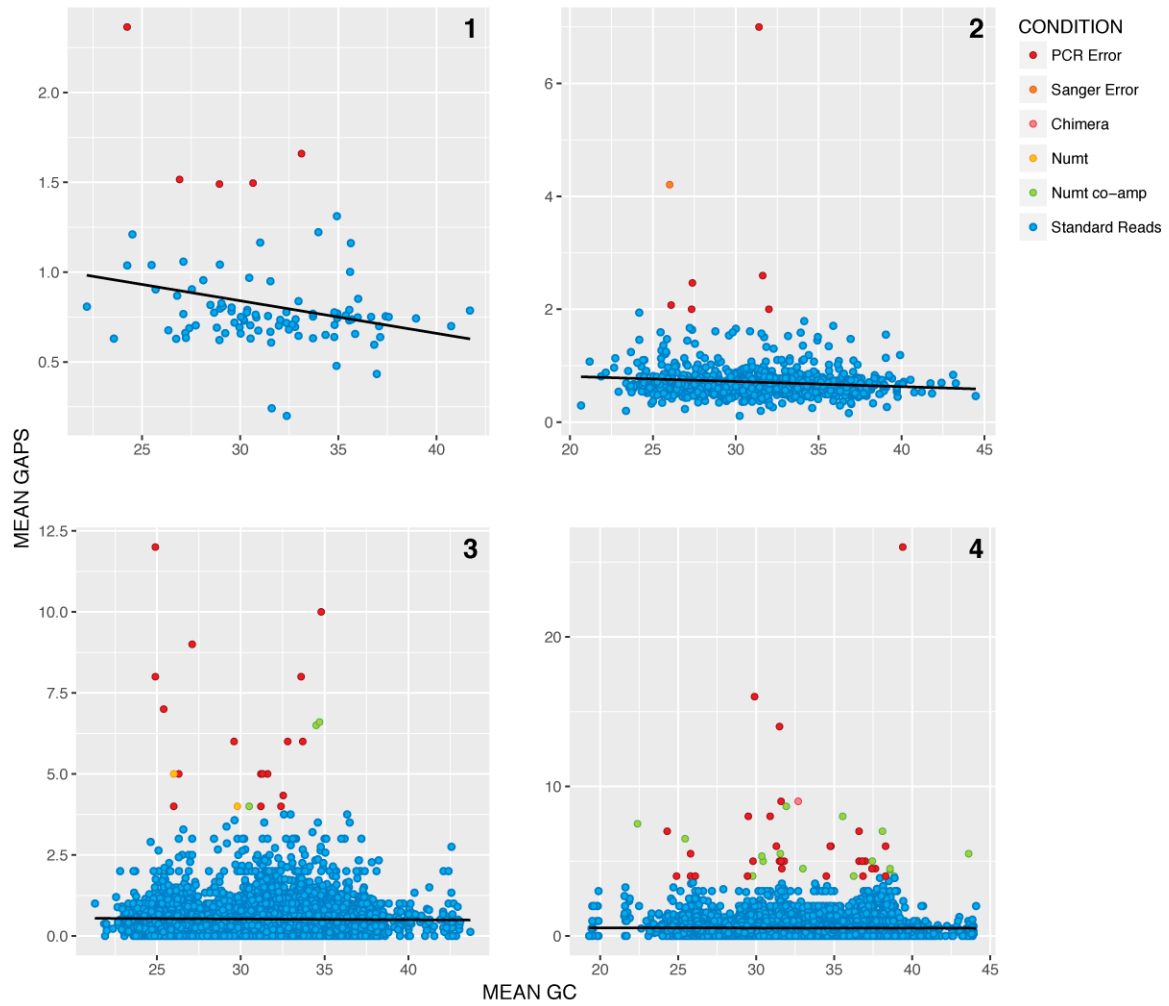


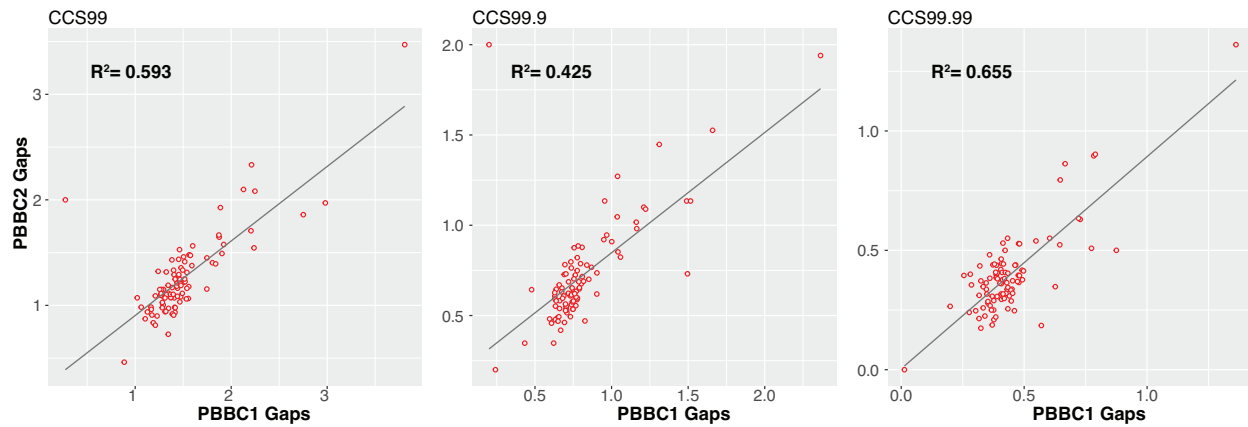
## Additional File 1



**Figure S1.** Sliding window (5 bp) analysis showing the distribution of homopolymer regions in the COI gene for the 95 taxa in library #1, the GC composition of the homopolymer regions, the incidence and position of indels in and substitutions in COI sequences for each window in the 99.9% SMRT partition. The incidence of substitutions and indels are shown per base pair.

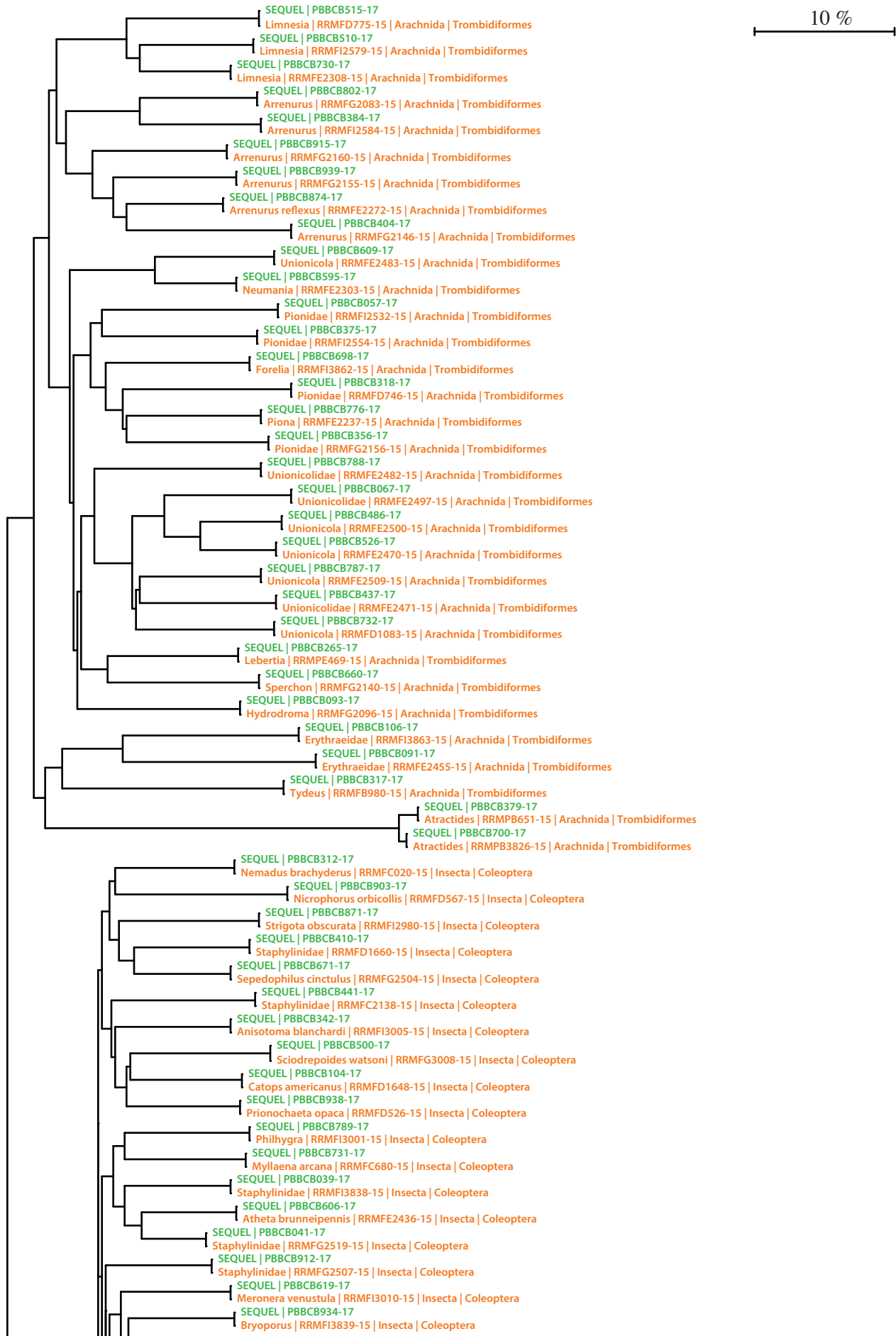


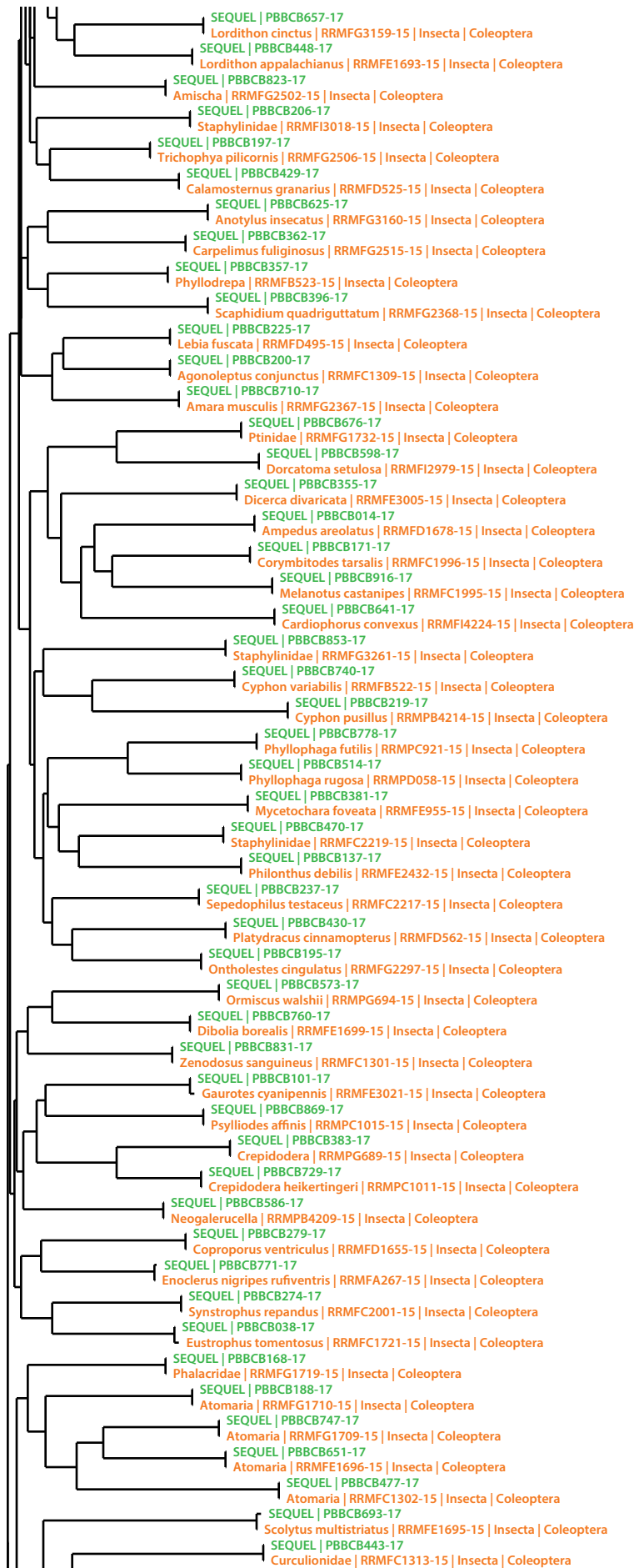
**Figure S2.** Bivariate plot showing the mean number of indels per CCS for each taxon in the four libraries versus the GC content of its COI amplicon. The number of indels is determined through comparison with the corresponding Sanger sequence for each taxon.

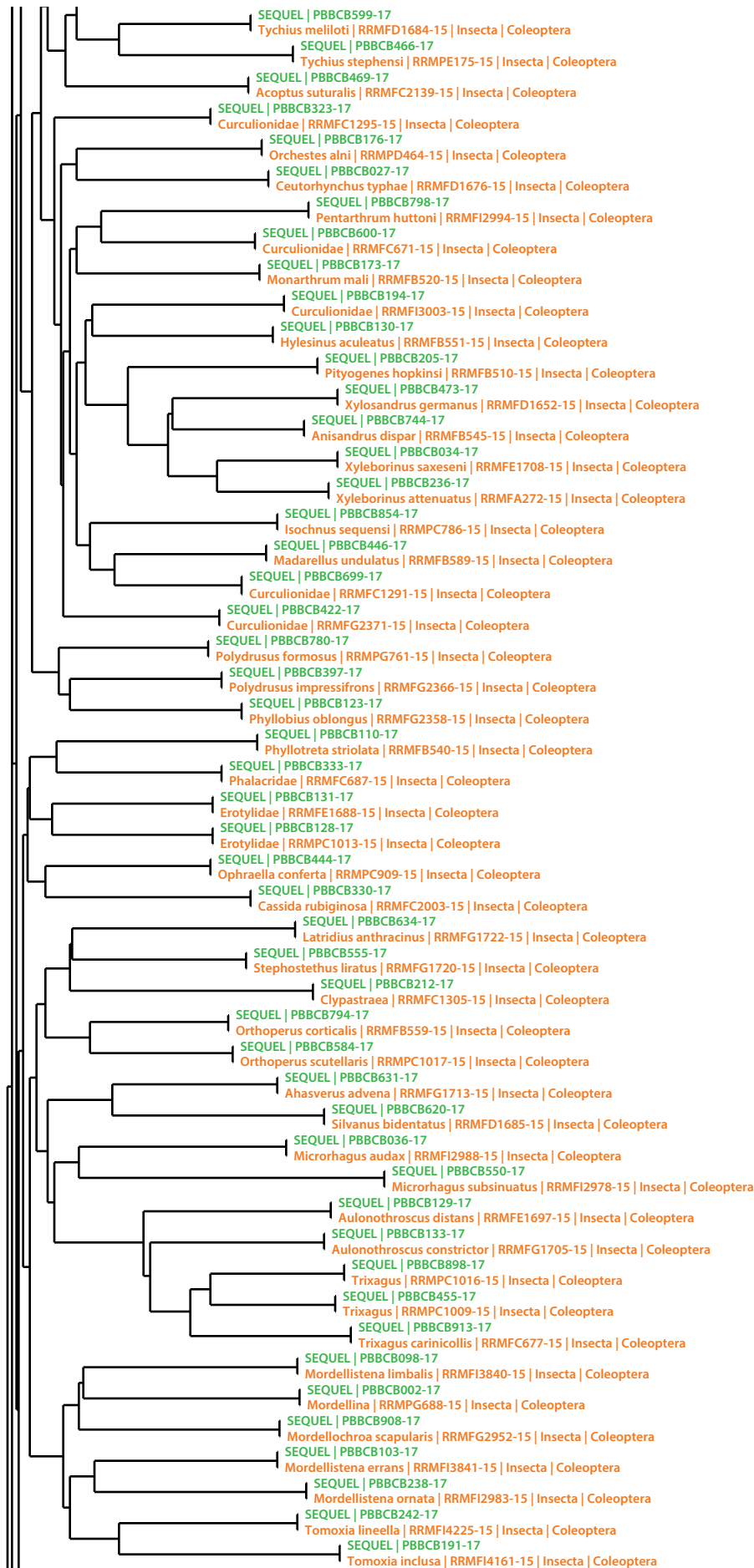


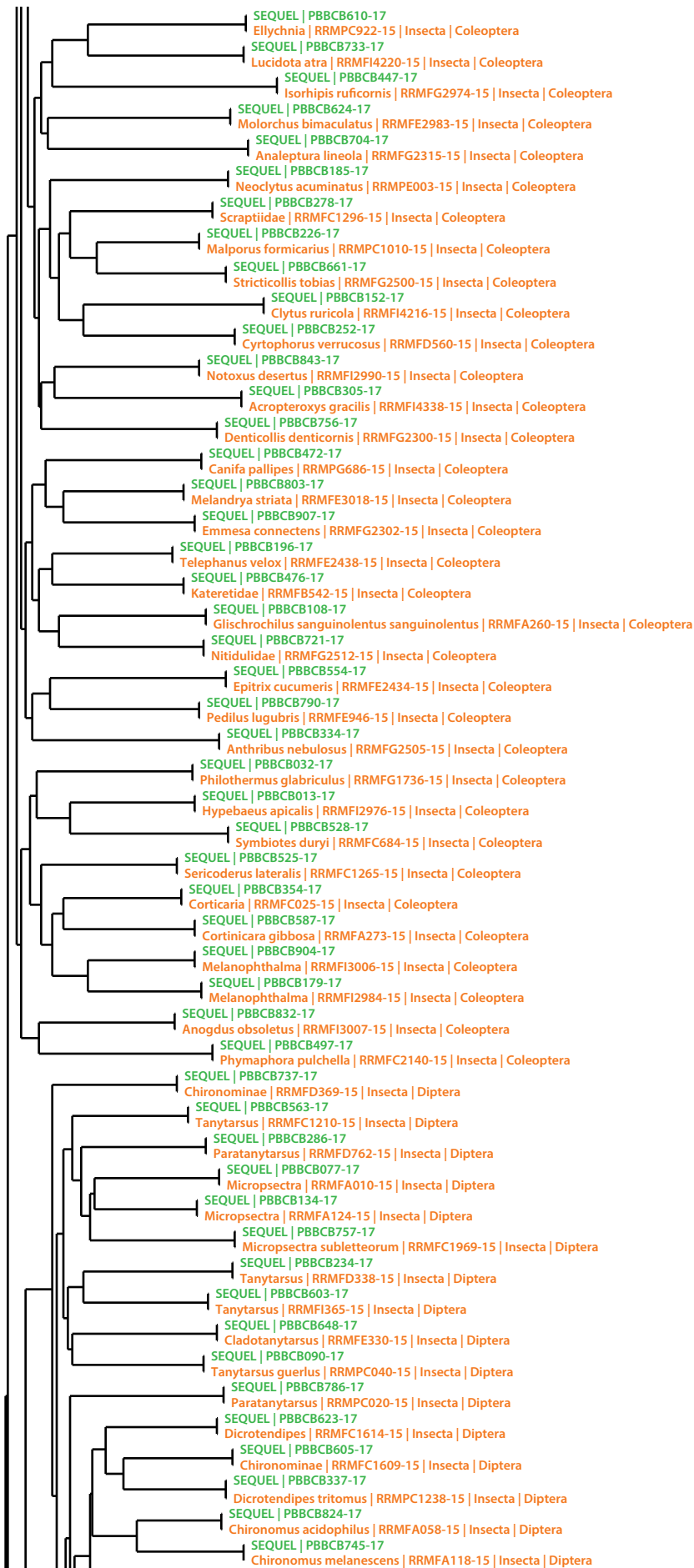
**Figure S3.** Bivariate plot showing the mean number of indels in SMRT sequences for the 95 specimens that were shared by Libraries #1 and #2.

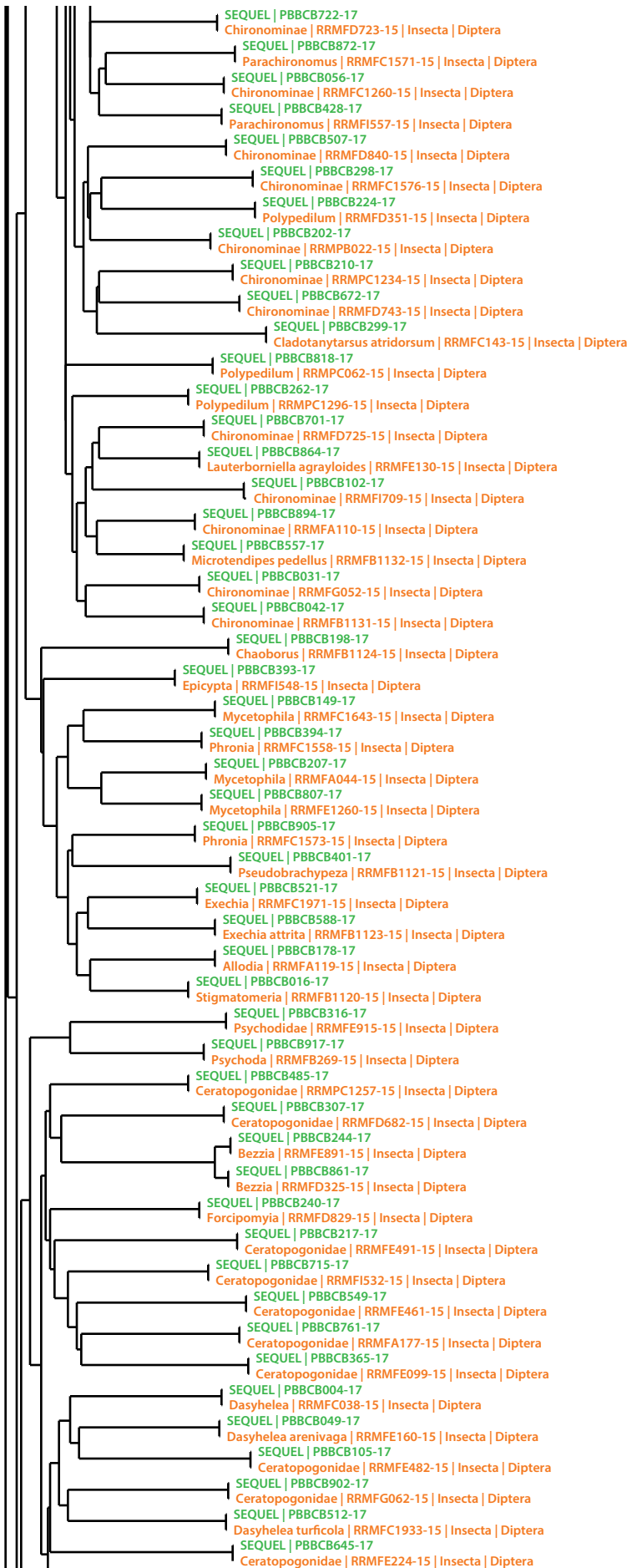
**Figure S4.** NJ tree of COI sequences for the 945 species in Library #2 showing the correspondence between Sanger and SMRT sequences. (19pp.)



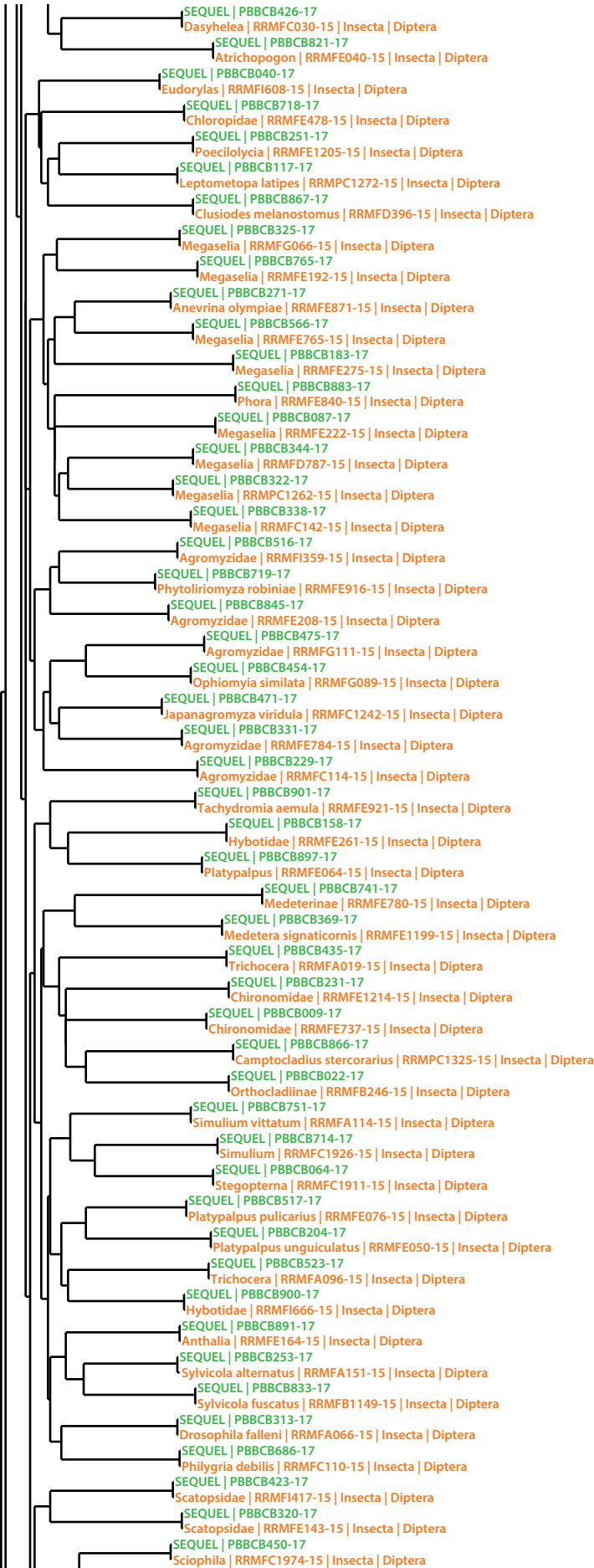


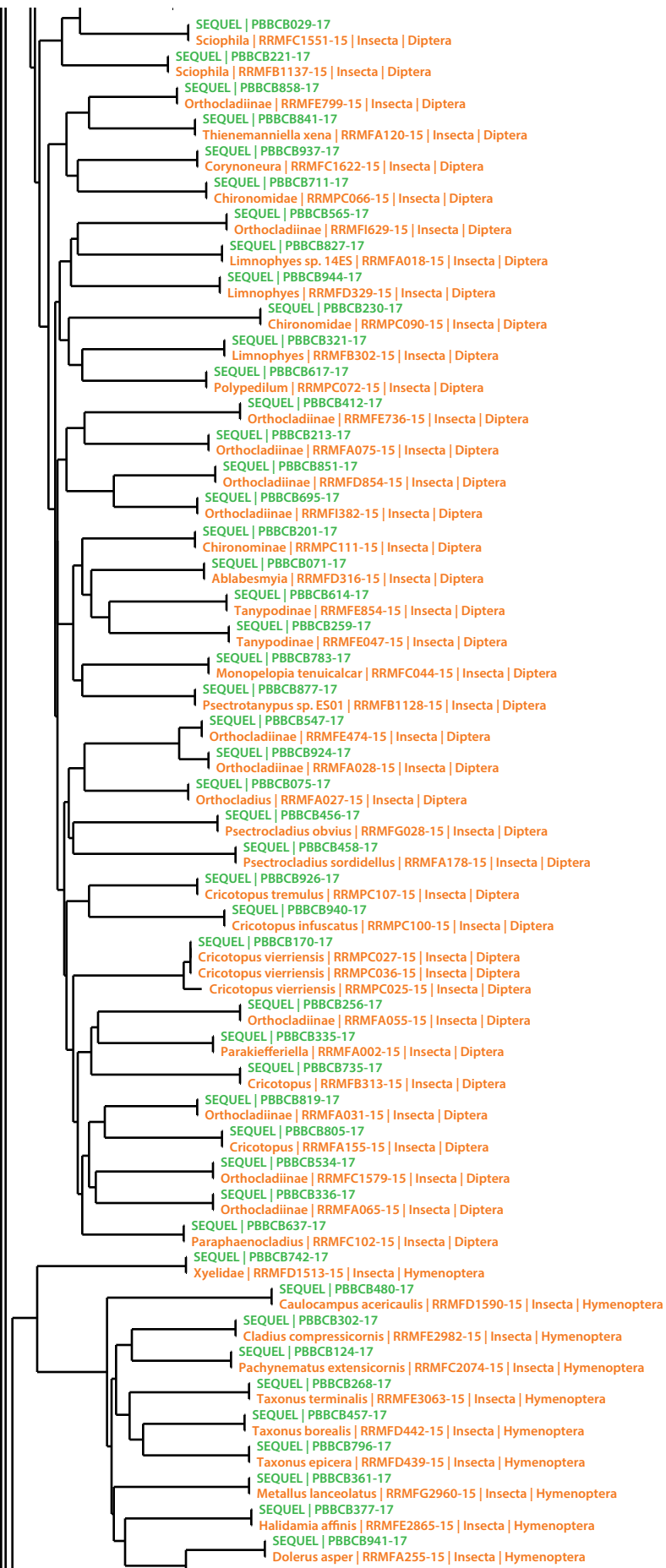


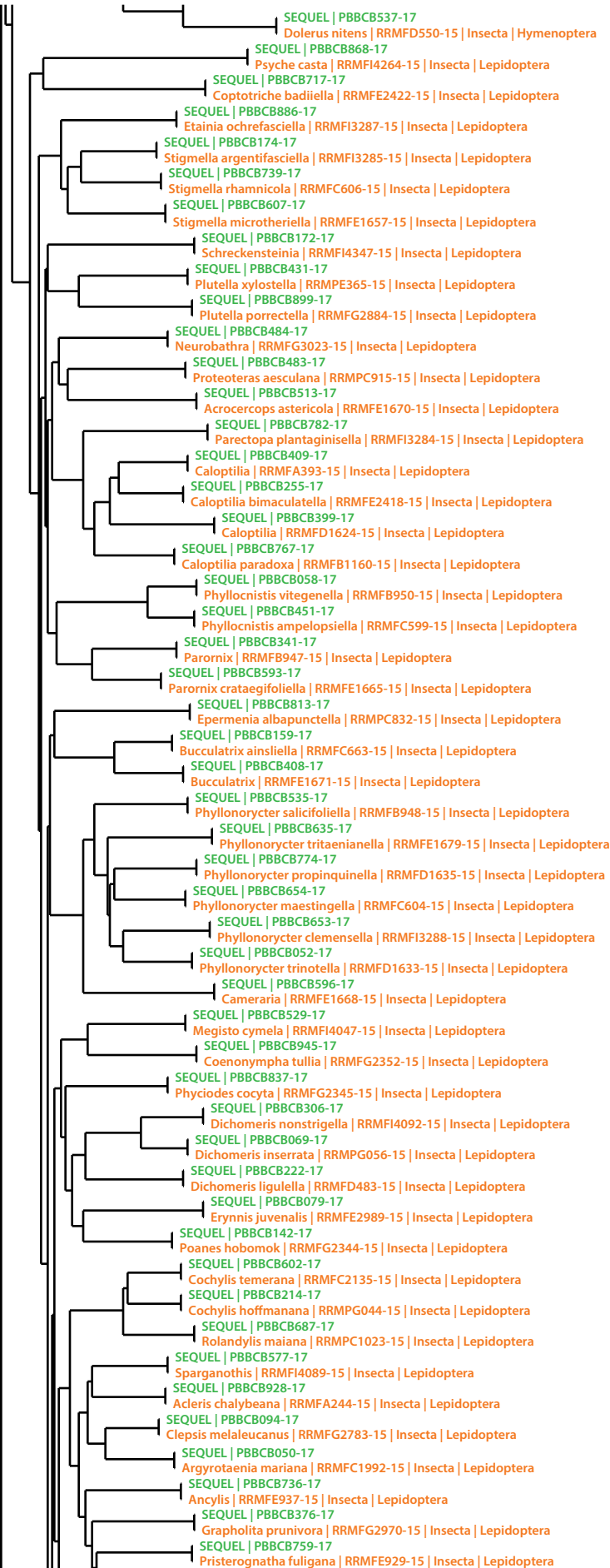


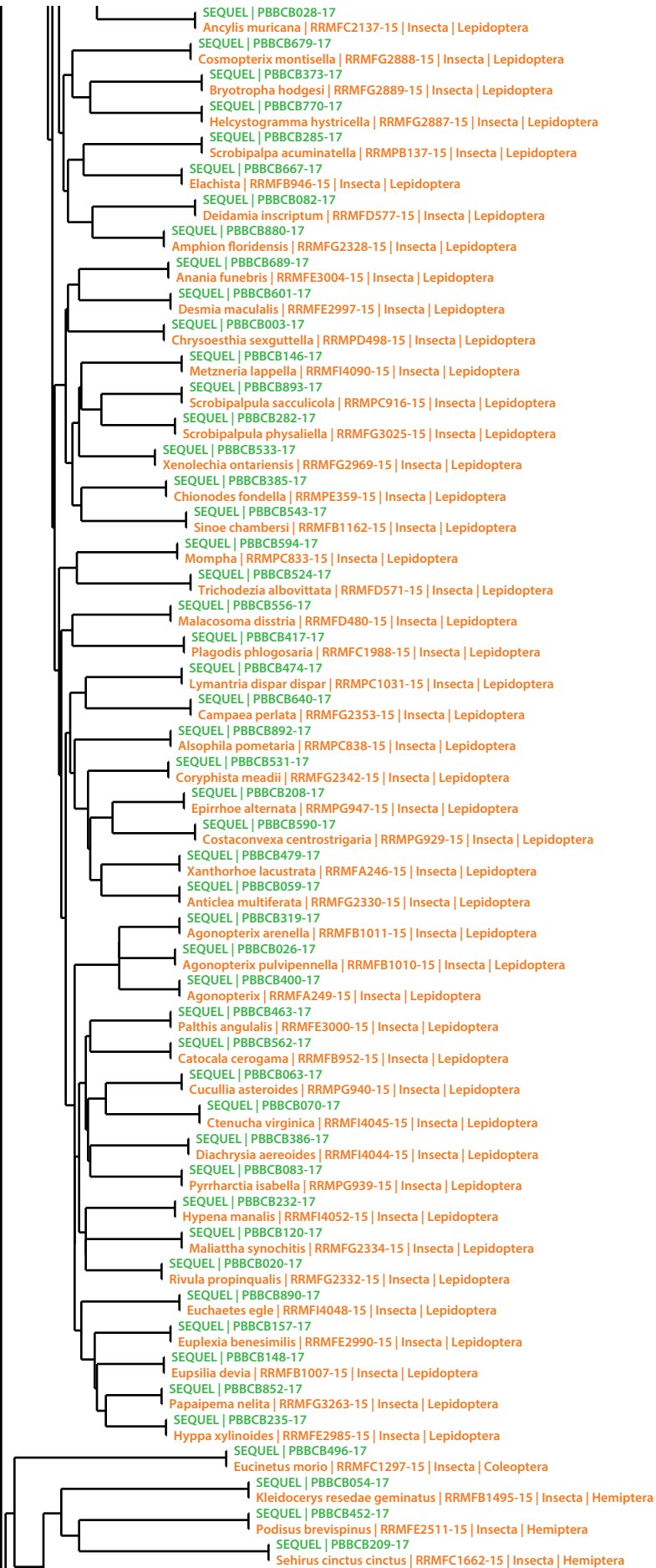


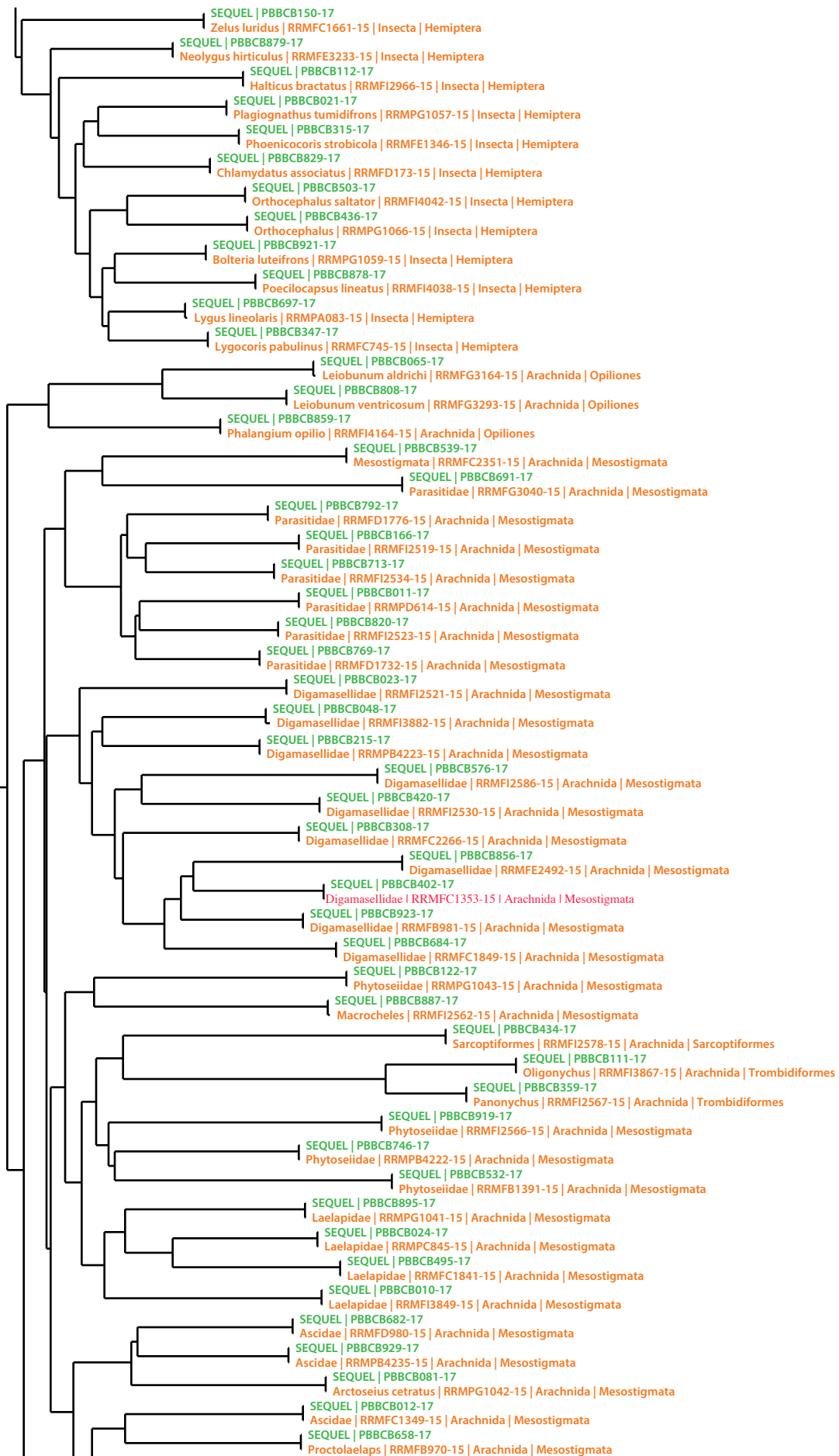


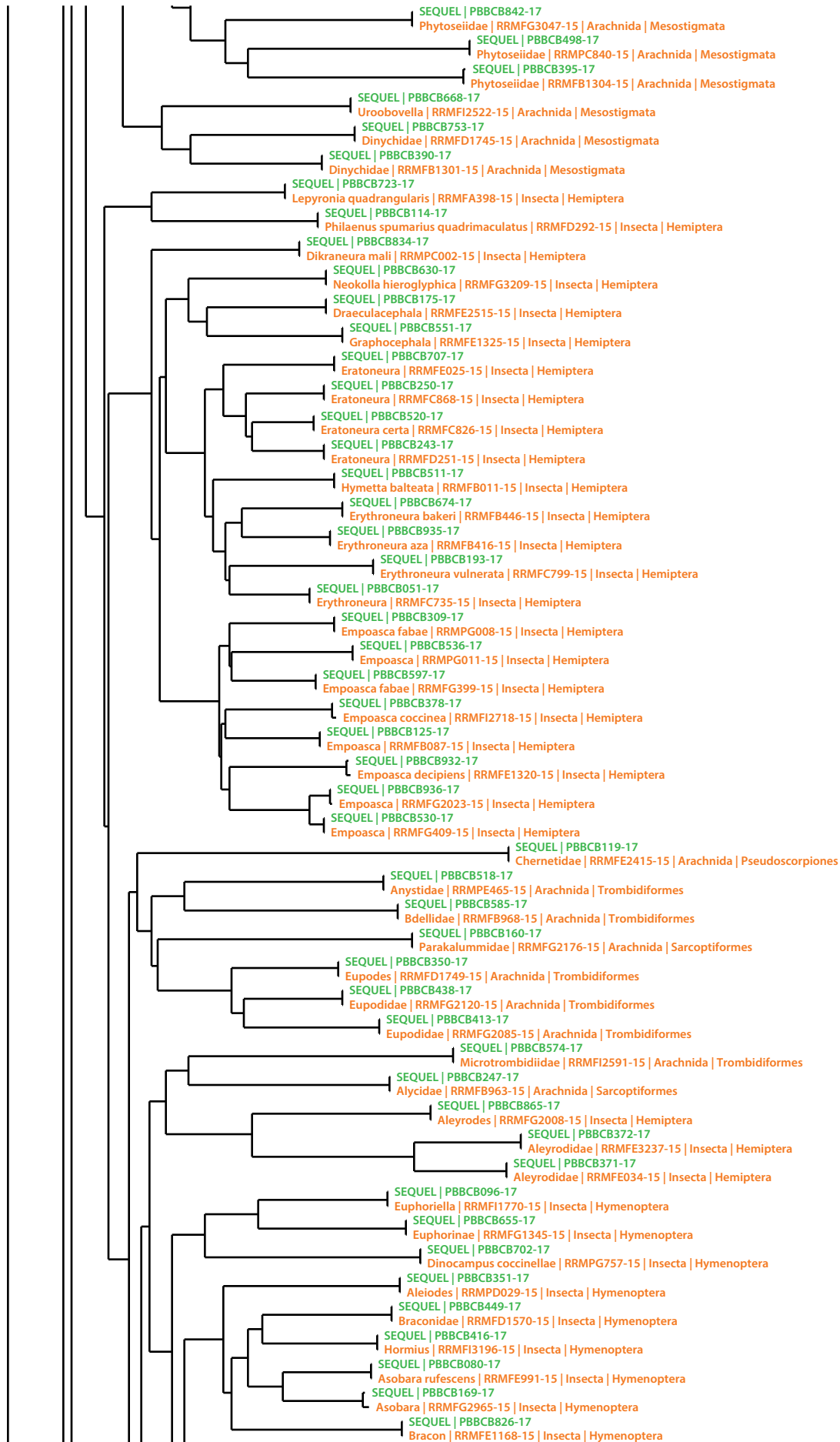


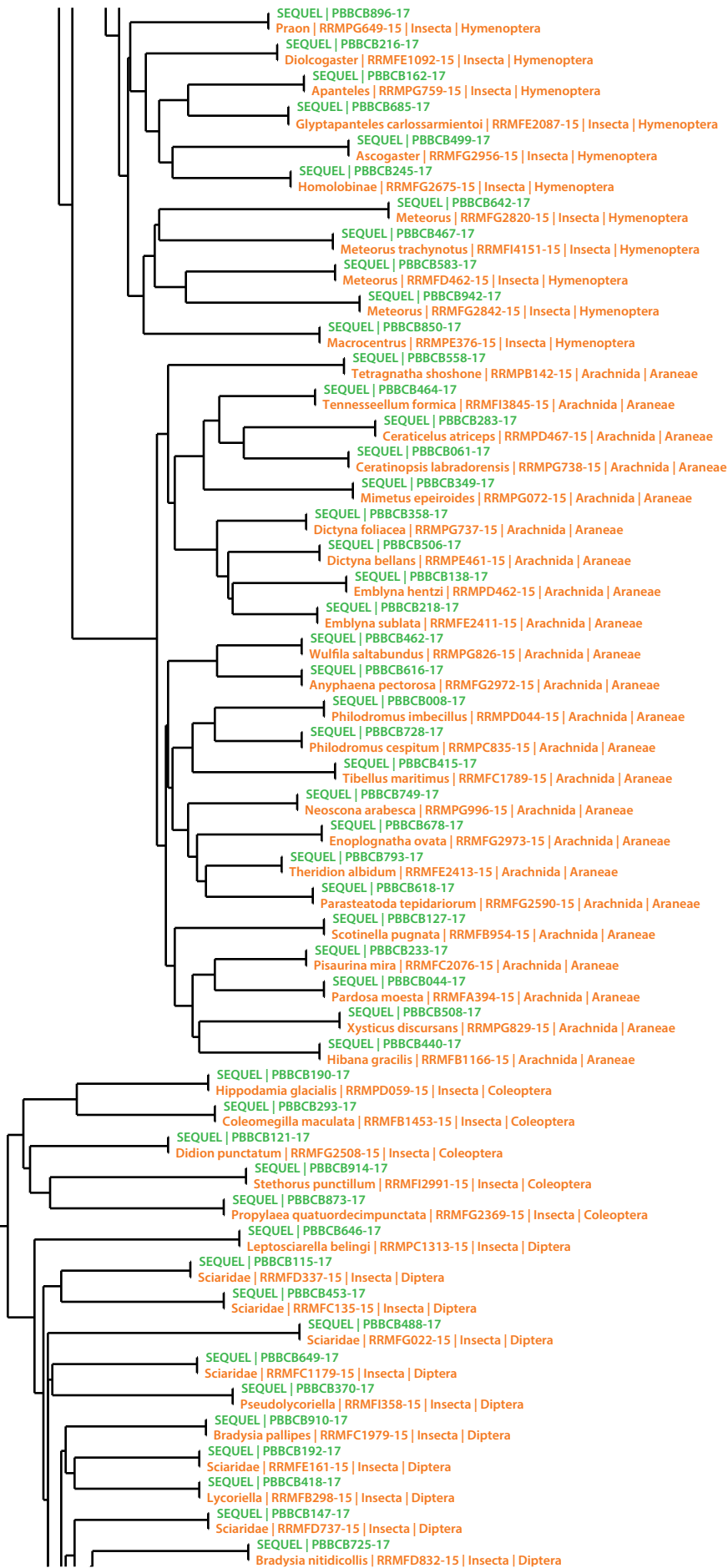


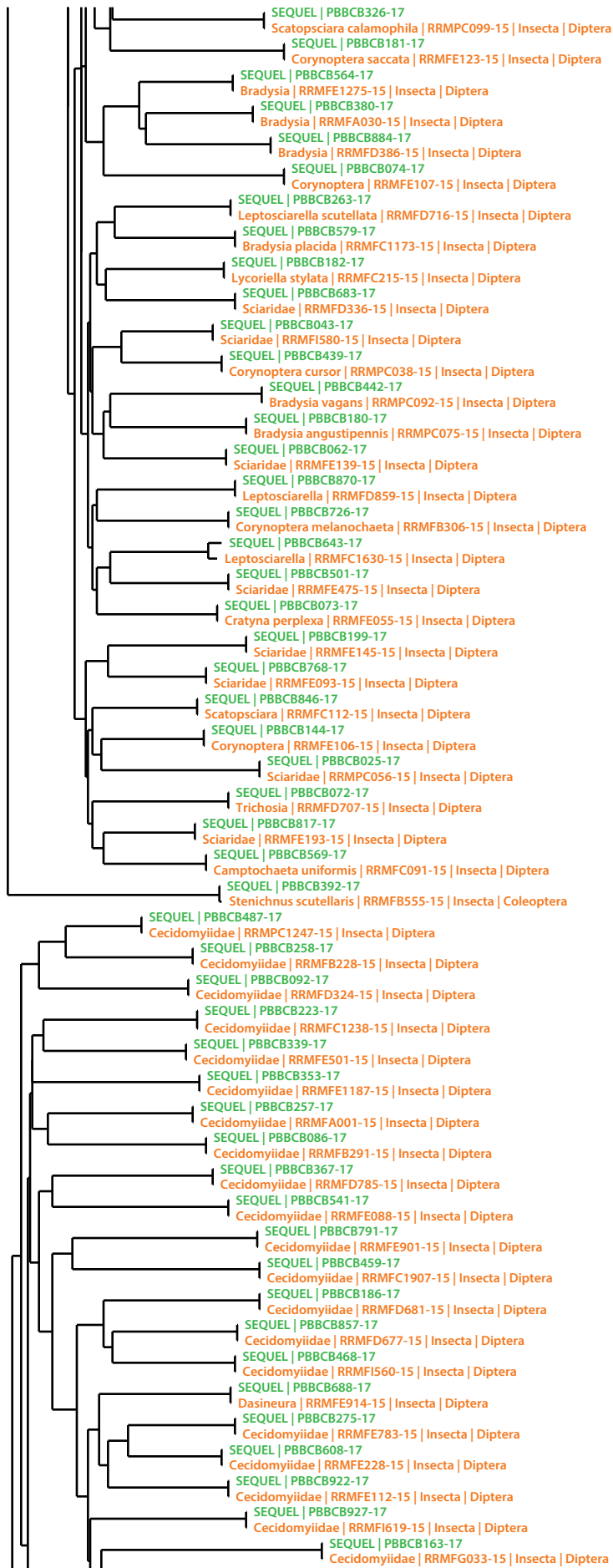




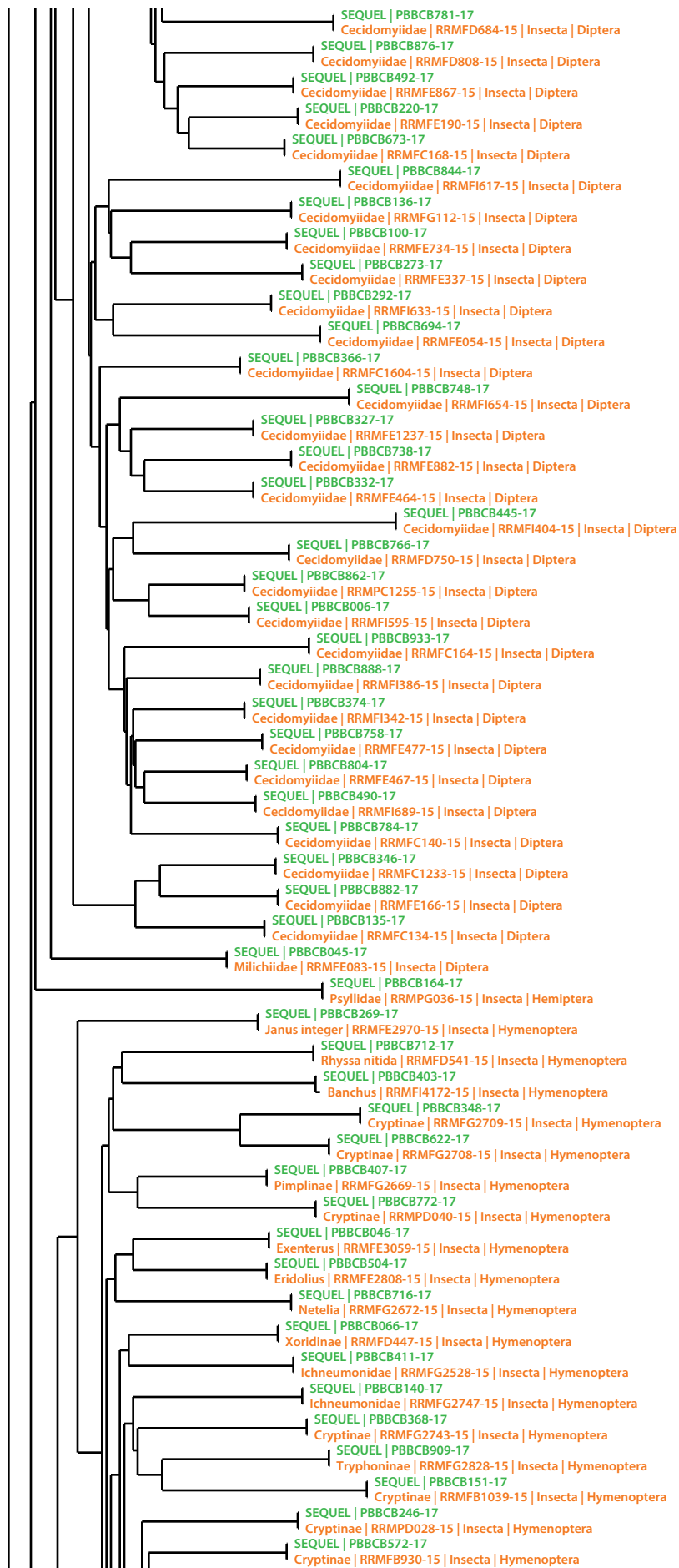


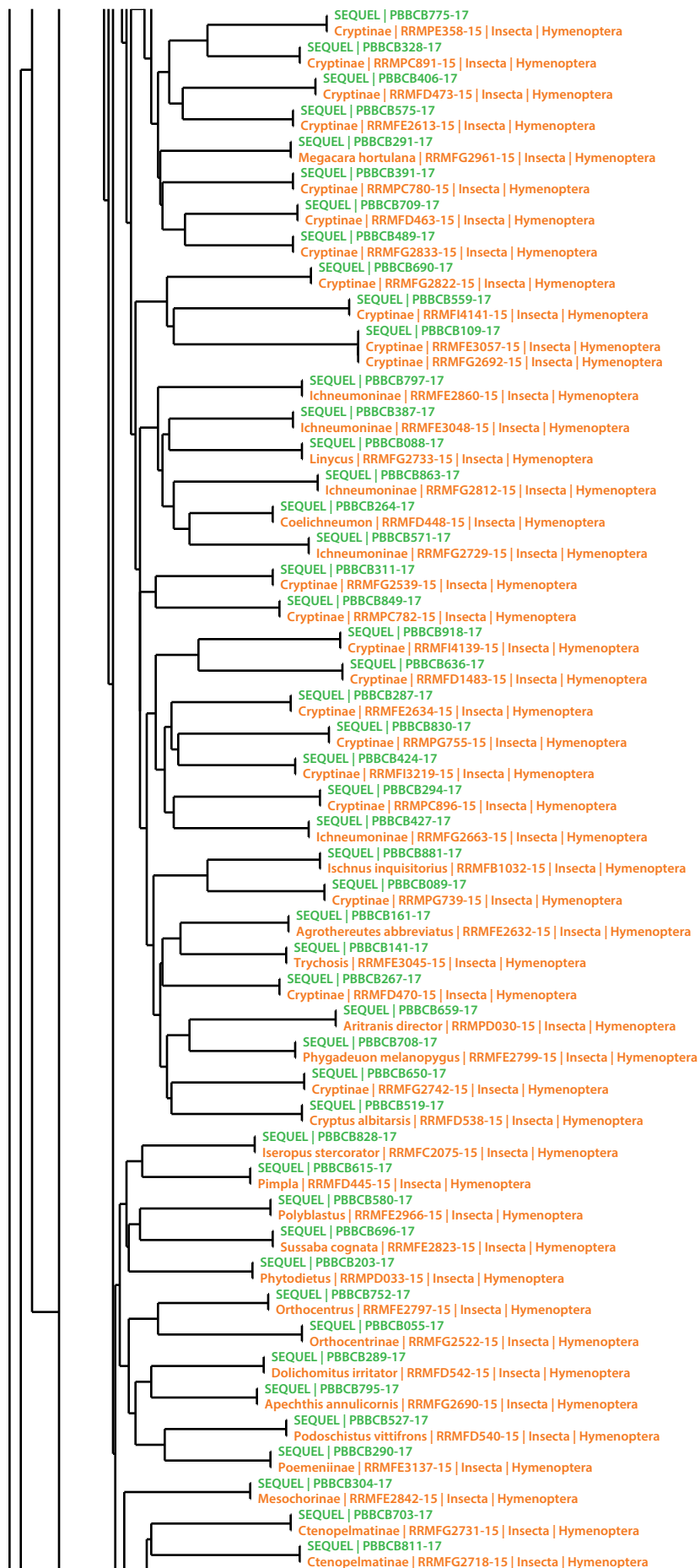


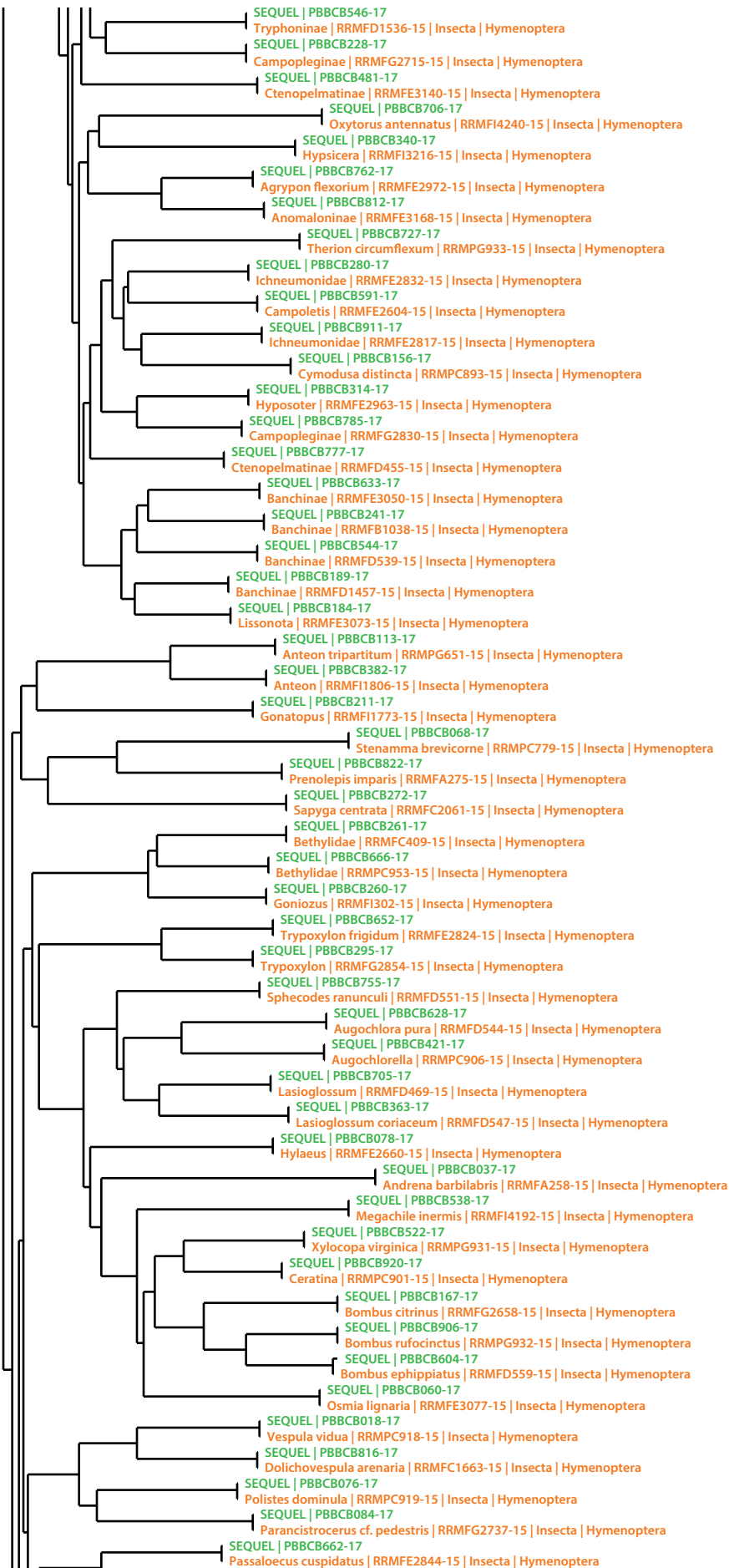


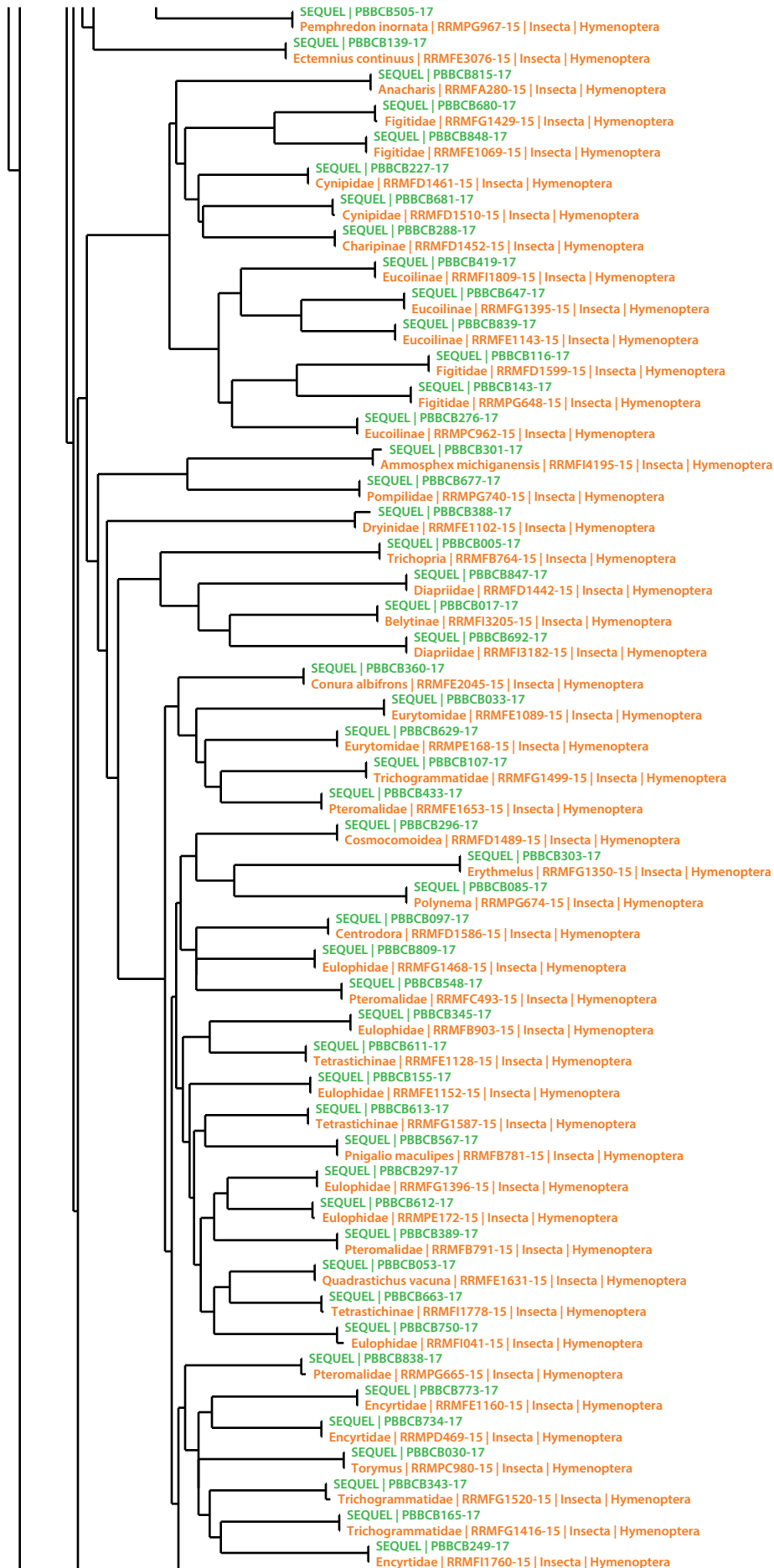


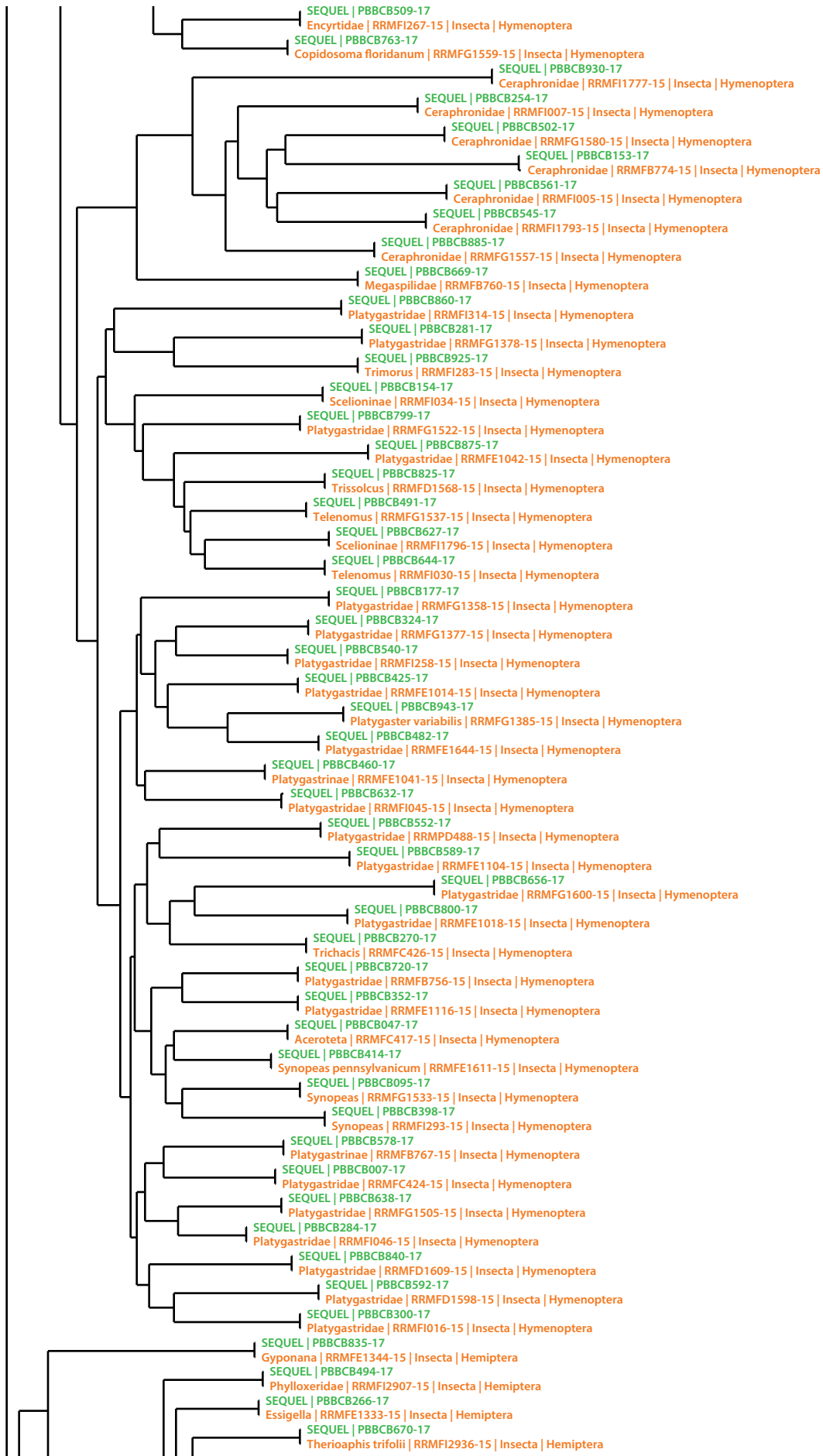


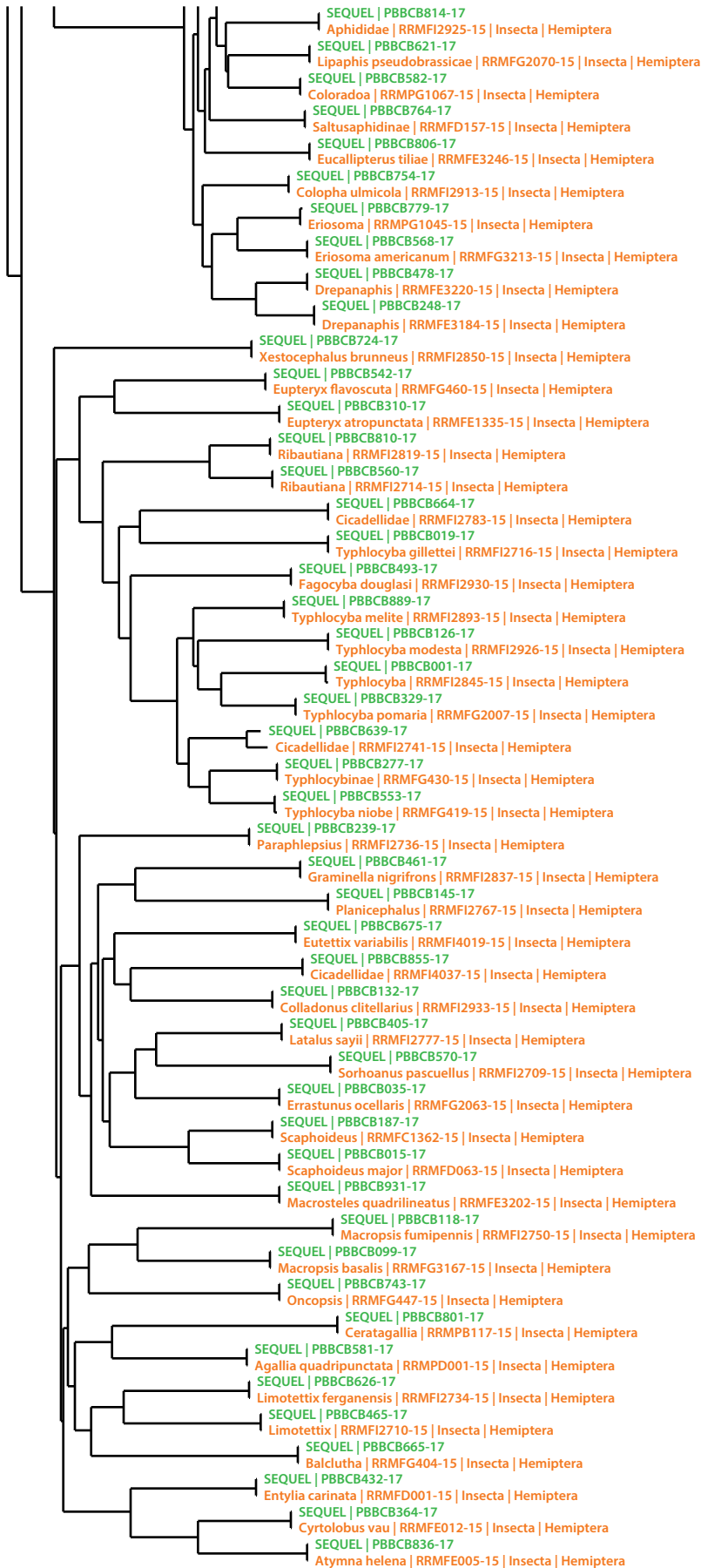


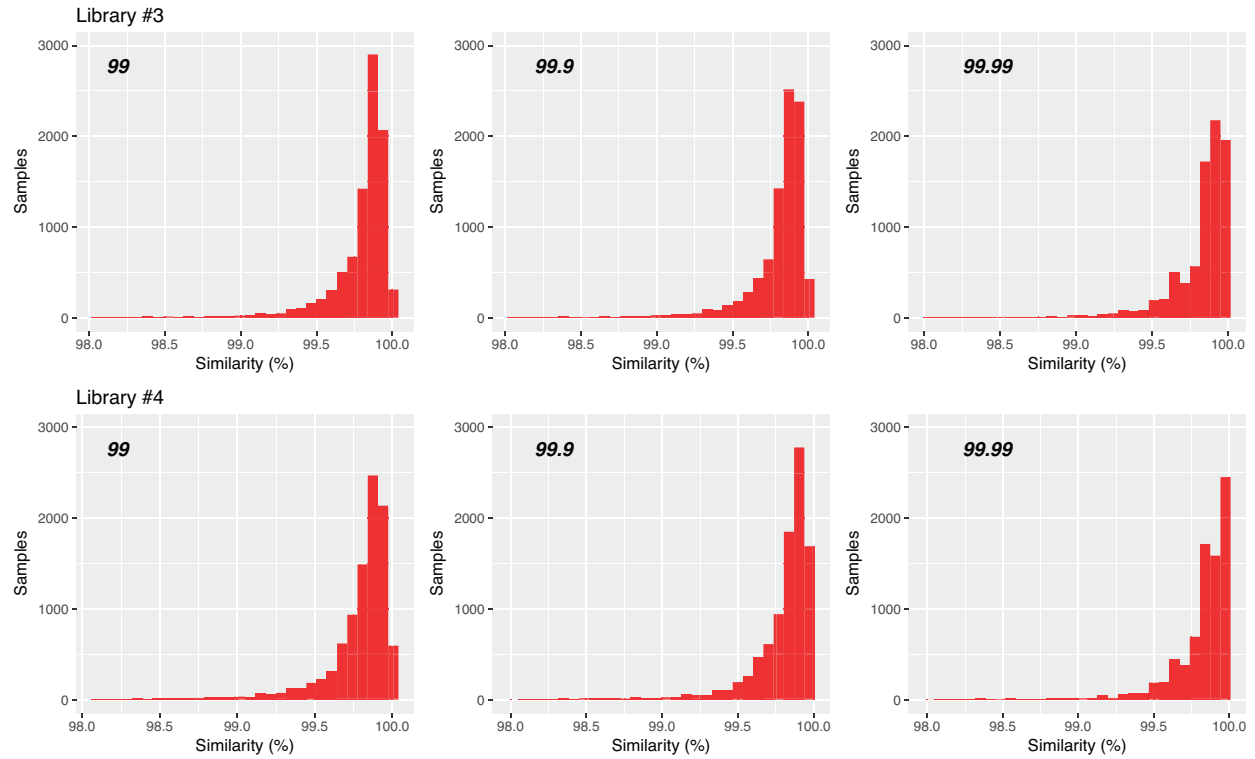




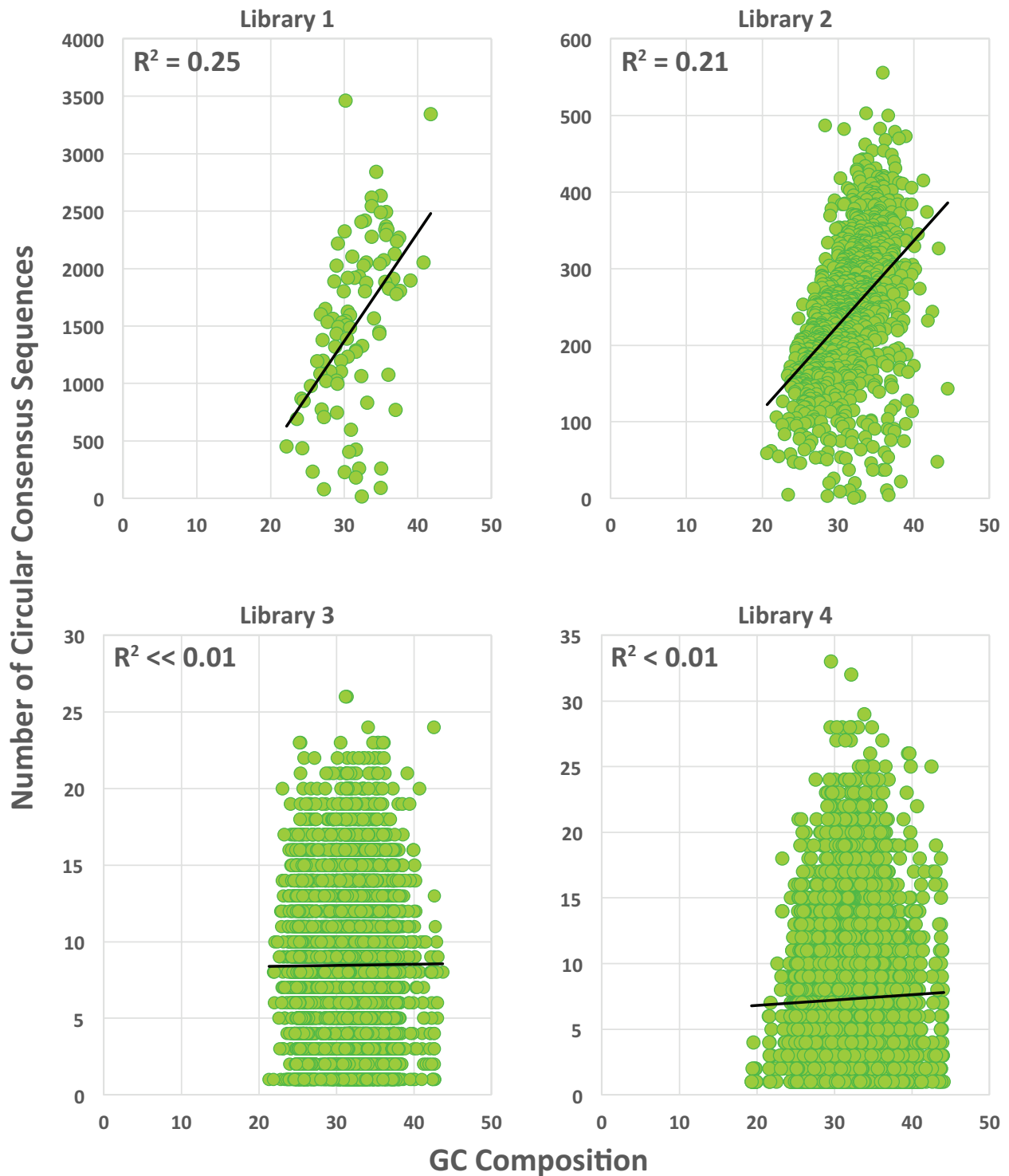








**Figure S5.** Distribution of similarity values between Sanger and SMRT sequences for the 9120 specimens in Library #3 and the 9830 specimens in Library #4 for all three SMRT partitions.



**Figure S6.** Relationship between the number of circular consensus sequences for a COI amplicon and its GC content. The 95 and 950 amplicon datasets (Libraries #1/2) were generated via the PCR1 protocol while the 9120 and 9830 amplicon datasets (Libraries #3/4) were generated via the PCR2 protocol.





**Figure S7.** Incidence of low quality Sanger reads in cases of target and non-target wells for Libraries #3/4.